

УДК 004.89+336.713

Т. О. Савчук, к. т. н., доц.; С. І. Петришин

ВИЗНАЧЕННЯ ЕВКЛІДОВОЇ ВІДСТАНІ МІЖ НАДЗВИЧАЙНИМИ СИТУАЦІЯМИ НА ЗАЛІЗНИЧНОМУ ТРАНСПОРТІ ПІД ЧАС КЛАСТЕРНОГО АНАЛІЗУ

У роботі проаналізовано можливість застосування технологій Data Mining під час аналізу надзвичайних ситуацій на залізничному транспорті. Формалізовано задачу кластерного аналізу, виявлено основні проблеми визначення відстані між надзвичайними ситуаціями під час такого аналізу. Визначено поняття звичайної, «зваженої» та квадрата евклідової відстані між надзвичайними ситуаціями на залізничному транспорті.

Ключові слова: евклідова відстань, надзвичайні ситуації, залізничний транспорт, кластерний аналіз, відстань, ступінь близькості, квадрат евклідової відстані, «зважена» евклідова відстань.

Вступ

У зв'язку зі збільшенням вантажоперевезень залізничним транспортом, враховуючи шкідливі та небезпечні вантажі, актуальними стають проблеми аналізу даних про надзвичайні ситуації, які можуть статися під час транспортування цих вантажів. При використанні стандартних математичних методів для аналізу таких ситуацій є ймовірність отримати недостовірні рішення, які при використанні системами підтримки прийняття рішень або аналітиками, збільшують ризик виконання недоцільних або помилкових дій ліквідаційними підрозділами. Тому достовірний аналіз з метою ідентифікації таких ситуацій є актуальною проблемою при розробці систем підтримки прийняття рішень для їх ліквідації / зменшення їх наслідків [1].

Технології аналізу даних, які ґрунтуються на застосуванні класичних статистичних підходів, мають низку недоліків під час використання для аналізу надзвичайних ситуацій на залізничному транспорті. Відповідні методи ґрунтуються на використанні усереднених показників, на підставі яких важко з'ясувати реальний стан такої ситуації. Методи математичної статистики виявилися корисними, насамперед, для перевірки заздалегідь сформульованих гіпотез та «грубого» розвідницького аналізу, що становить основу оперативної аналітичної обробки даних.

Окрім того, стандартні статистичні методи не враховують нетипові спостереження, що є недопустимим при аналізі надзвичайних ситуацій на залізниці. Проте окремі нетипові значення можуть бути важливими для дослідження, характеризуючи виняткові явища. При цьому сама ідентифікація цих спостережень та їх подальший аналіз і детальний розгляд є корисними для розуміння сутності досліджуваної надзвичайної ситуації на залізничному транспорті. Як показують сучасні дослідження, саме такі події можуть стати вирішальними щодо майбутньої поведінки та розвитку надзвичайної ситуації [1].

При аналізі надзвичайних ситуацій, які можуть виникнути під час перевезення небезпечних вантажів, необхідно оперувати реальними значеннями, організувати пошук неявних закономірностей у даних, самостійну побудову гіпотез про взаємозв'язки параметрів і характеристик таких ситуацій.

Огляд існуючих методів аналізу надзвичайних ситуацій на залізничному транспорті

Велику кількість задач, зокрема, і задач аналізу надзвичайних ситуацій на залізничному транспорті, допомагають розв'язати технології Data Mining у залежності від характеру розв'язку задач (описові задачі та задачі прогнозування), як показано на рис. 1. Всі алгоритми аналізу даних поділяють на supervised learning (навчання з учителем) та

unsupervised learning (навчання без учителя) (рис. 1). У першому випадку задача аналізу розв'язується в декілька етапів. Спочатку за допомогою певного алгоритму будується модель даних, які аналізують. Потім ця модель навчається на навчальних вибірках до того моменту, поки вона не почне працювати коректно. Unsupervised learning використовують, коли немає ніяких попередніх знань про аналізовані дані. Основними із задач Data Mining є такі: класифікація, регресія, пошук асоціативних правил і кластеризація (табл. 1) [2]. Розглянемо їх на прикладі аналізу надзвичайних ситуацій на залізниці.

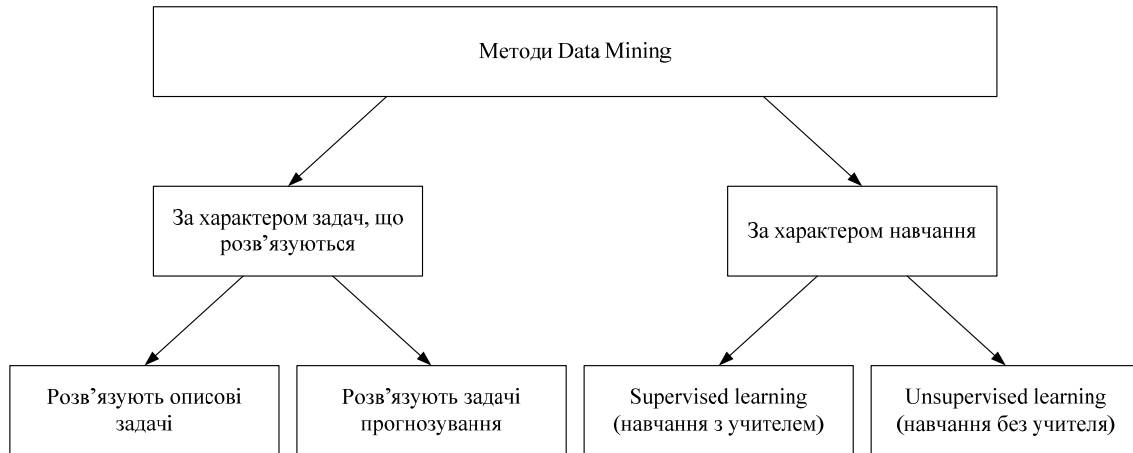


Рис. 1. Класифікація методів, які використовують в технологіях Data Mining для аналізу надзвичайних ситуацій на залізничному транспорті

Задача класифікації зводиться до визначення класу надзвичайної ситуації на залізничному транспорті за її характеристиками. У цій задачі множина класів надзвичайних ситуацій, до яких може бути віднесений об'єкт дослідження, наперед відома.

Використання класифікації під час аналізу надзвичайних ситуацій є простим у застосуванні та характеризується наявністю великої кількості ефективних підходів до розв'язання цієї задачі, але це не є актуальним у досліджуваних питаннях. Крім того, слід відзначити, що недоліки використання класифікації у цьому випадку такі [2]:

- 1) потужність навчальної вибірки має бути достатньо великою;
- 2) у навчальну вибірку мають входити надзвичайні ситуації, які представляють всі класи, які є проблемним при аналізі таких ситуацій;
- 3) для кожного класу повинна бути достатньо потужна множина надзвичайних ситуацій у навчальній вибірці, що важко отримати під час аналізу надзвичайних ситуацій на залізниці;
- 4) проблема *overfitting*, сутність якої полягає в тому, що класифікаційна функція добре адаптується до даних, і, якщо серед них зустрічаються помилки, або аномальні значення, то функція інтерпретує їх як частину внутрішньої структури даних, що є неприйнятним для аналізу надзвичайних ситуацій на залізничному транспорті;
- 5) проблема *underfitting*, яка полягає в тому, що під час перевірки класифікатора виявляється велика кількість помилок, що є неприйнятним для предметної області, яку аналізують.

Задача регресії як і задача класифікації дозволяє за відомими характеристиками надзвичайної ситуації на залізничному транспорті визначити значення певного його параметра. На відміну від класифікації значенням цього параметра є не множина класів, а множина дійсних чисел, що є неактуальним в аналізі надзвичайних ситуацій на залізничному транспорті.

Таблиця 1

Характеристика методів Data Mining, які можливо використовувати під час аналізу надзвичайних ситуацій на залізничному транспорті

Назва задачі	Сутність задачі	Переваги задачі	Недоліки задачі
Класифікація	Визначення класу надзвичайної ситуації на залізничному транспорті за її відомими характеристиками	Простота у застосуванні; наявність великої кількості ефективних підходів до розв'язання цієї задачі	Потужність навчальної вибірки має бути достатньо великою; у навчальну вибірку мають входити надзвичайні ситуації, які представляють всі класи; для кожного класу повинна бути достатньо потужна множина надзвичайних ситуацій у навчальній вибірці; проблема overfitting; проблема underfitting
Регресія	Дозволяє за відомими характеристиками надзвичайної ситуації на залізничному транспорті, визначити значення певного їх параметра	Простота у застосуванні; наявність великої кількості підходів до розв'язання цієї задачі	Неможливість розв'язання задачі ідентифікації надзвичайних ситуацій на залізничному транспорті
Пошук асоціативних правил	Визначення залежностей, які часто повторюються серед надзвичайних ситуацій	Можливість знаходження певних закономірностей між надзвичайними ситуаціями; легке сприйняття людиною правил; проста інтерпретація мов програмування	Правила не завжди корисні, оскільки є три види асоціативних правил: корисні, тривіальні, незрозумілі
Кластеризація	Пошук незалежних кластерів у множині даних щодо аналізованих надзвичайних ситуацій на залізничному транспорті	Ітераційний пошук оптимального результату; можливість використання методів утворення кластерів, вибору інформативних ознак та мір близькості між двома об'єктами, об'єктом і кластером, двома кластерами; побудова науково обґрунтованої класифікації багатовимірних спостережень на підставі сукупності відібраних показників та виявлення внутрішніх зв'язків між надзвичайними ситуаціями на залізниці, які аналізуємо	Визначення на вході кількості ітерацій при пошуку рішення

Метою пошуку асоціативних правил є визначення залежностей, які часто повторюються серед надзвичайних ситуацій. Знайдені залежності подано у вигляді правил і можуть бути використані як для кращого розуміння природи даних, які аналізуються, так і для прогнозування виникнення певних подій [2].

Перевагами пошуку асоціативних правил є те, що вони дозволяють знаходити певні закономірності між надзвичайними ситуаціями, що є актуальним для поданої предметної області. А також не є важливим при аналізі таких ситуацій легке сприйняття людиною правил та проста інтерпретація мов програмування.

Недоліками пошуку асоціативних правил є те, що правила, які знаходять у результаті Наукові праці ВНТУ, 2010, № 3

такого аналізу не завжди корисні, оскільки є три види асоціативних правил: корисні, тривіальні, незрозумілі. Така ситуація є неприйнятною для аналізу надзвичайних ситуацій на залізниці.

Задача кластеризації полягає в пошуку незалежних кластерів у множині даних про надзвичайні ситуації, які аналізуємо на залізничному транспорті. Це дозволяє зрозуміти дані. Крім того, групування однорідних даних дозволяє зменшити їх кількість для спрощення аналізу в подальшому [2].

Перевагами кластеризації є ітераційний пошук оптимального результату, що підвищує ймовірність знаходження оптимального рішення; можливість використання методів утворення кластерів і вибору інформативних ознак та мір близькості між двома об'єктами, об'єктом і кластером, двома кластерами, що є актуальним при аналізі надзвичайних ситуацій на залізниці; побудова науково обґрунтованої класифікації багатовимірних спостережень на підставі сукупності відібраних показників та виявлення внутрішніх зв'язків між надзвичайними ситуаціями на залізниці, які аналізуємо.

Складність реалізації кластеризації регулюється визначенням на вході кількості ітерацій при пошуку рішення, що є важливим для предметної області, яку аналізуємо оскільки таким чином визначається точність прогнозованих результатів роботи алгоритму ідентифікації надзвичайної ситуації, яка підлягає аналізу і стану її розвитку.

Отже, за результатами проведеного аналізу можна зробити висновок про доцільність використання кластеризації для аналізу надзвичайних ситуацій на залізничному транспорті. Ці дані свідчать про те, що кластеризація характеризується ітераційним пошуком оптимального рішення, можливістю вибору інформативних ознак та мір близькості двома об'єктами, об'єктом і кластером, двома кластерами, побудовою науково обґрунтованої класифікації багатовимірних спостережень на підставі сукупності відібраних показників та виявлення внутрішніх зв'язків між надзвичайними ситуаціями на залізниці, які аналізуються.

Постановка проблеми

Кластерний аналіз являє собою спосіб групування багатовимірних об'єктів, якими є надзвичайні ситуації на залізничному транспорті. Поданий аналіз засновано на поданні результатів окремих надзвичайних ситуацій точками певного геометричного простору з подальшим виділенням груп цих точок (кластерів, таксонів). Кластерний аналіз дозволяє виділяти компактні, віддалені одна від одної групи надзвичайних ситуацій, припускаючи «природне» розбиття множини на зони скупчення таких ситуацій, що дозволяє стверджувати про однорідність дій при ліквідації надзвичайних ситуацій, які належать до одного кластера. Кластерний аналіз використовують для аналізу надзвичайних ситуацій в таких випадках [3]:

– дані про надзвичайні ситуації подано у вигляді матриці близькості або відстаней між конкретними такими ситуаціями (1):

$$\Psi = \begin{Bmatrix} 0 & \psi_{12} & \dots & \psi_{1n} \\ \psi_{21} & 0 & \dots & \psi_{2n} \\ \dots & \dots & \dots & \dots \\ \psi_{n1} & \psi_{n2} & \dots & 0 \end{Bmatrix}, \quad (1)$$

де ψ_{ij} – відстань між параметрами векторів $\psi(Y_i, Y_j)$, де Y_i та Y_j – надзвичайні ситуації на залізничному транспорті;

n – кількість надзвичайних ситуацій на залізничному транспорті, інформацію про які потрібно опрацьовувати (потужність бази даних надзвичайних ситуацій на залізничному транспорті).

– дані про надзвичайні ситуації подано у вигляді точок у багатовимірному просторі (2):

$$Y = \{Y_1, Y_2, \dots, Y_n\} = \begin{Bmatrix} y_{11} & y_{12} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{nm} \end{Bmatrix}, \quad (2)$$

де Y_i – конкретна надзвичайна ситуація на залізничному транспорті;

y_{ij} – значення конкретного j -го параметру i -ої надзвичайної ситуації;

m – кількість параметрів надзвичайних ситуацій, які збережено в базі даних.

При аналізі надзвичайних ситуацій можна використовувати два наведені способи подання даних. Для цього потрібно опрацьовувати потужні бази даних, тому у цьому випадку другий спосіб (2) є доцільнішим, оскільки матриця близькості при збільшенні бази даних надзвичайних ситуацій на залізничному транспорті на одну таку ситуацію збільшується на $((l+1)^2 - l^2)$ елементів, де l – початкова кількість надзвичайних ситуацій на залізничному транспорті в базі даних. Також недоліком першого способу подання даних є те, що під час додавання нової надзвичайної ситуації в базу даних потрібно враховувати (використовуючи обчислювальну техніку або без неї) її ступені близькості або відстані до кожної збереженої в базі даних надзвичайної ситуації. Та особі, яка приймає рішення щодо ліквідації надзвичайної ситуації невідомо координати кожної надзвичайної ситуації на залізничному транспорті. Щодо другого способу, то тут під час додавання в базу даних нової надзвичайної ситуації кількість її елементів збільшиться на m . Також перевагою другого способу подання даних є можливість визначити координати кожної надзвичайної ситуації як геометричного об'єкта.

Отже, задачу кластерного аналізу надзвичайних ситуацій на залізничному транспорті можна сформулювати в такий спосіб.

Нехай існує множина надзвичайних ситуацій на залізничному транспорті $Y = \{Y_i\} (i = \overline{1, n})$. Кожна з надзвичайних ситуацій має m характеристик. Потрібно розбити статичний m -вимірний діапазон зміни значень аналізованих ознак надзвичайних ситуацій на інтервали групування. Тобто, множину Y розбити на $k (k \leq n)$ кластерів таким чином, щоб конкретна надзвичайна Y_i ситуація належала до одного і тільки одного кластера. Основною умовою при цьому є максимальна подібність надзвичайних ситуацій, які належать до одного кластера, і максимальна несхожість надзвичайних ситуацій з різних кластерів [3].

Відстані та ступені близькості надзвичайних ситуацій на залізничному транспорті

Складності у формалізації задачі кластерного аналізу надзвичайних ситуацій на залізничному транспорті пов'язані з визначенням поняття їх однорідності [4] та слабкою структурованістю даних.

Загалом однорідність двох i -ої та j -ої надзвичайної ситуацій на залізничному транспорті визначається завданням правила обрахунку величини ψ_{ij} , що характеризує або відстань $a(Y_i, Y_j)$ між об'єктами Y_i та Y_j із досліджуваної множини надзвичайних ситуацій на залізничному транспорті $Y = \{Y_i\} (i = \overline{1, n})$, або ступінь близькості $\omega(Y_i, Y_j)$ між тими ж ситуаціями. Якщо задана функція $a(Y_i, Y_j)$, то близькі за значенням цієї метрики надзвичайні ситуації вважаються однорідними, тобто такими, що належать до одного кластера. Але при цьому необхідно порівнювати $a(Y_i, Y_j)$ з певними пороговими значеннями, які визначаються в будь-якому випадку. Означений підхід доцільно використовувати для визначенні міри близькості $\omega(Y_i, Y_j)$ при формуванні однорідних кластерів надзвичайних ситуацій на залізничному транспорті. При цьому потрібно дотримуватися таких вимог:

- вимоги симетрії ($\omega(Y_i, Y_j) = \omega(Y_j, Y_i)$);
- вимоги максимальної подібності надзвичайних ситуацій ($\omega(Y_i, Y_i) = \max(\omega(Y_i, Y_j))$);
- вимоги відповідності відстані між надзвичайними ситуаціями на залізничному транспорті та мірою близькості між ними (якщо $a(Y_1, Y_2) \geq a(Y_2, Y_3)$ то $\omega(Y_1, Y_2) \leq \omega(Y_2, Y_3)$).

Вимірювання відстані між надзвичайними ситуаціями

Відстанню між надзвичайними ситуаціями Y_i та Y_j або метрикою називається невід'ємна дійсна функція $a(Y_i, Y_j)$, якщо [5]:

- $a(Y_i, Y_j) \geq 0$ для всіх Y_i та Y_j з множини $Y = \{Y_i\} (i = \overline{1, n})$;
- $a(Y_i, Y_j) = 0$ тоді і тільки тоді, коли $Y_i = Y_j$;
- $a(Y_i, Y_j) = a(Y_j, Y_i)$;
- $a(Y_i, Y_j) \leq a(Y_i, Y_k) + a(Y_k, Y_j)$, де Y_i, Y_j та Y_k – будь-які три надзвичайні ситуації на залізничному транспорті з множини $Y = \{Y_i\} (i = \overline{1, n})$.

При кластерному аналізі надзвичайних ситуацій виникає проблема вимірювання відстані між окремими ситуаціями. Основні труднощі, які виникають при цьому такі [5]:

- неоднозначність вибору способу нормування;
- неоднозначність визначення відстані між об'єктами.

Якщо розглянути результати досліджень деяких надзвичайних ситуацій на залізничному транспорті, то на рис. 2 за результатами цих досліджень побудуємо кореляційне поле. Масштабність за осями обирається довільно. На рис. 2(а) виділено певні класи А, Б, В, що при зміні масштабу вісі «Температура» (рис. 2(б)) змінюються – формуються клас A_1 , який ідентичний попередній візуалізації, та клас B_1 , який об'єднує класи Б і В, які є недопустимим під час аналізу надзвичайних ситуацій на залізниці. Визначити відстань між надзвичайними ситуаціями в цьому випадку неможливо, оскільки ознаки виміряні в різних одиницях. Потрібно нормувати показники, тобто перевести їх в безвимірні величини з метою коректного вимірювання відстані між надзвичайними ситуаціями на залізничному транспорті. Нормуванням є перехід до певного однакового опису всіх ознак, до введення нової умовної одиниці вимірювання, що допускати формальне порівняння надзвичайних ситуацій на залізничному транспорті.

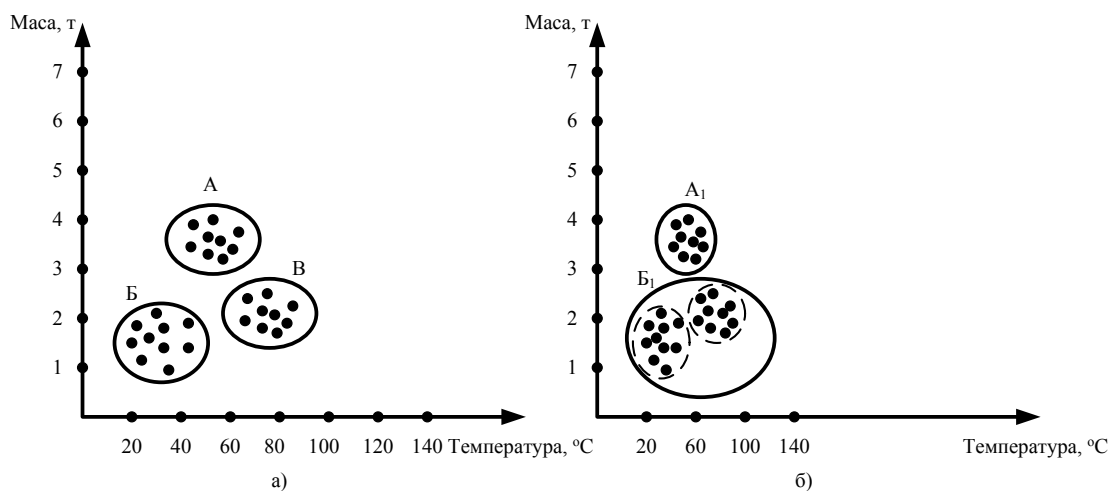


Рис. 2. Розбиття сукупності надзвичайних ситуацій на залізничному транспорті на кластери в залежності від масштабів вимірювання змінних

Використання евклідової відстані під час аналізу надзвичайних ситуацій на залізничному транспорті

Евклідова відстань [6] є одним з найбільш використовуваних метрик у кластерному аналізі, оскільки вона відповідає інтуїтивним уявленням про близькість і своєю квадратичною формою відповідає класичним статистичним конструкціям. Геометрично подану метрику доцільно використовувати для об'єднання об'єктів у скупченнях, які є типовими для слабко корельованих множин.

Формула загальної евклідової відстані має вигляд (3):

$$a_E(Y_i, Y_j) = \sqrt{(y_{i1} - y_{j1})^2 + (y_{i2} - y_{j2})^2 + \dots + (y_{im} - y_{jm})^2}, \quad (3)$$

де $a_E(Y_i, Y_j)$ – евклідова відстань між двома надзвичайними ситуаціями на залізничному транспорті Y_i та Y_j ;

$y_{i1}, y_{i2}, \dots, y_{im}$ – вектор значень характеристик, який описує i -ту надзвичайну ситуацію на залізничному транспорті;

$y_{j1}, y_{j2}, \dots, y_{jm}$ – вектор значень характеристик, який описує j -ту надзвичайну ситуацію на залізничному транспорті.

Подану метрику доцільно застосовувати в таких випадках:

- значення параметрів $y_{i1}, y_{i2}, \dots, y_{im}$ однорідні за своїм фізичним змістом, і якщо встановлено, що всі вони однаково важливі з точки зору розв'язку задачі про віднесення надзвичайної ситуації на залізничному транспорті до певного кластера;

- простір ознак збігається з геометричним простором дійсності і поняття близькості надзвичайних ситуацій збігається з поняттям геометричної близькості в цьому просторі.

Отже, подану метрику можливо застосовувати при аналізі надзвичайних ситуацій на залізничному транспорті у випадках, коли будуть аналізуватись такі характеристики, які будуть близькі за своїм фізичним змістом, що є неприйнятним для поданої предметної області, оскільки потрібно аналізувати всі фактори для отримання достовірного результату.

Для того, щоб краще розрізняти віддалені об'єкти інколи використовують квадрат евклідової відстані [6] (4):

$$a_E(Y_i, Y_j)^2 = (y_{i1} - y_{j1})^2 + (y_{i2} - y_{j2})^2 + \dots + (y_{im} - y_{jm})^2. \quad (4)$$

При необхідності визначення важливості кожної характеристики λ_l надзвичайної ситуації на залізничному транспорті (наприклад, температури в цистерні при перевезенні легкозаймистих речовин та відносна вологість повітря), яка буде пропорційною ступеню її важливості з точки зору віднесення певної надзвичайної ситуації до конкретного кластера, доцільно використовувати «зважену» евклідову відстань [4] (5):

$$a_{3E}(Y_i, Y_j) = \sqrt{\lambda_1 \cdot (y_{i1} - y_{j1})^2 + \lambda_2 \cdot (y_{i2} - y_{j2})^2 + \dots + \lambda_m \cdot (y_{im} - y_{jm})^2}, \quad (5)$$

де $a_{3E}(Y_i, Y_j)$ – «зважена» евклідова відстань між двома надзвичайними ситуаціями на залізничному транспорті Y_i та Y_j ;

$\lambda_1, \lambda_2, \dots, \lambda_m$ ($0 \leq \lambda_l \leq 1 (l = \overline{1, m})$) – вектор значень вагових коефіцієнтів, які відповідають характеристикам y_1, y_2, \dots, y_m надзвичайних ситуацій на залізничному транспорті;

$y_{i1}, y_{i2}, \dots, y_{im}$ – вектор значень характеристик, який описує i -ту надзвичайну ситуацію на залізничному транспорті;

$y_{j1}, y_{j2}, \dots, y_{jm}$ – вектор значень характеристик, який описує j -ту надзвичайну ситуацію на залізничному транспорті.

Для визначення вектора значень вагових коефіцієнтів $\lambda_1, \lambda_2, \dots, \lambda_m$ використовують

навчальні вибірки надзвичайних ситуацій на залізничному транспорті або досвід експертів. Спроби визначити вагових коефіцієнтів $\lambda_1, \lambda_2, \dots, \lambda_m$ лише за інформацією, яка міститься у вхідних даних, не дають потрібного результату і можуть збільшити похибку отриманого результату.

Отже, подана метрика є прийнятною для проведення кластерного аналізу надзвичайних ситуацій на залізничному транспорті, оскільки тут враховується значимість кожної характеристики надзвичайної ситуації на залізничному транспорті, що підвищує достовірність результату.

Висновок

Отже, для аналізу надзвичайних ситуацій на залізничному транспорті доцільно використовувати кластеризацію, яка характеризується ітераційним пошуком оптимального рішення; можливістю вибору інформативних ознак та мір близькості між двома об'єктами, об'єктом і кластером, двома кластерами; побудовою науково обґрунтованої класифікації багатовимірних спостережень на підставі сукупності відібраних показників та виявлення внутрішніх зв'язків між надзвичайними ситуаціями на залізниці, які аналізують.

Серед розглянутих метрик доцільною для використання при аналізі надзвичайних ситуацій на залізниці є «зважена» евклідова відстань як така, що враховує значимість кожної характеристики надзвичайної ситуації на залізничному транспорті, що сприяє отриманню достовірного результату.

СПИСОК ЛІТЕРАТУРИ

1. Савчук Т. О. Використання ієрархічних методів кластеризації для аналізу надзвичайних ситуацій на залізничному транспорті / Т. О. Савчук, С. І. Петришин // Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах» (м. Хмельницький) – 2009. – №1 – С. 193 – 198.
2. Барсегян А. А. Методы и модели анализа данных: OLAP и Data Mining. / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод – СПб.: БХВ-Петербург, 2004. – 336 с.
3. Савчук Т. О. Порівняльний аналіз використання методів кластеризації для ідентифікації надзвичайних ситуацій на залізничному транспорті / Т. О. Савчук, С. І. Петришин // Наукові праці Донецького національного технічного університету. – Серія «Інформатика, кібернетика і обчислювальна техніка». – 2010. – Випуск 11(134). – С. 135 – 141.
4. Айвазян С. А. Прикладная статистика: Классификация и снижение размерности. / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков. – М.: Финансы и статистика, 1989. – 607 с.
5. Мандель И. Д. Кластерный анализ. / И. Д. Мандель. – М.: Финансы и статистика, 1988. – 176 с.
6. Дюран Б. Кластерный анализ / Б. Дюран, П. Одел.; пер. с англ. – М.: Статистика, 1977. – 128 с.

Савчук Тамара Олександрівна – к. т. н., доцент кафедри комп'ютерних наук.

Петришин Сергій Іванович – магістрант кафедри комп'ютерних наук.

Вінницький національний технічний університет.